

リンケージマッピング

三木哲郎・叶 林・荻原俊男

ヒト全ゲノム上の 5000 個以上のマイクロサテライト多型を用い、CEPH の大家系を連鎖分析することにより、精密度の高いリンケージマッピングが完成した。その遺伝子座位は、STS として登録され、YAC クローンの隣接地図を作成するときなどの物理的地図作成に役だっている。すべてのマイクロサテライトマーカーと各マーカー間の遺伝子距離の情報は、インターネットの経由で、Genethon, GDB, CHLC などから入手可能となった。詳細なリンケージマッピングは、全ゲノム解析の基準点として役だっている。

Key words 【リンケージマッピング】【マイクロサテライト多型】
【STS (sequence-tagged site)】

はじめに リンケージマッピングの実際の研究は、CEPH (Centre d'Etude du Polymorphisme Humain = Human Polymorphism Study Center) 家系を利用した国際的な共同作業により以前より存在したが、マイクロサテライト多型の有用性が見いだされ応用されるようになってから、飛躍的に進んだ分野である。現在、5000 個以上のマイクロサテライト多型が STS (sequence-tagged site) としてマッピングされている。これらが均等に分布すると仮定すれば、平均 600 kb に 1 個のマイクロサテライト多型が存在することになる。それぞれのマイクロサテライト多型に対応する YAC クローンが判明しており、またヘテロ接合性の高いマーカーが選択され、ゲノム上を一定の間隔で並んだマーカーのセットが市販されている。ここでは、多型性遺伝子マーカーの定義、種類、実際の連鎖分析について概説する。

多型性遺伝子マーカーは、血液型の多型や酵素のアイソザイムなどの蛋白質マーカーから、10 数年前、DNA

マーカーに交代した。さらに、DNA マーカーは、RFLP (restriction fragment length polymorphism), VNTR (variable number of tandem repeat) から、対立遺伝子が多く、PCR で簡単にタイピングできるマイクロサテライト多型へと移行している¹⁾。1996 年、フランスの Genethon グループは、全ゲノム上で 5264 個の CA 反復配列からなるマイクロサテライト多型について、CEPH 家系を使った連鎖分析によりその遺伝子座位を決定し、STS として登録した²⁾。STS は、全ゲノムの塩基配列決定の基準点となる。別に、このマイクロサテライト多型による STS (genetic map) と RH (radiation hybrid) mapping と STS-content map³⁾ の 3 種により、最終的に 15,086 個の STS が報告された。この数は、平均 199 kb に 1 個の STS が存在することになる³⁾。マイクロサテライト多型は、遺伝子地図作成と遺伝性疾患の解析に最も役だつ遺伝子マーカーとなった。

1. 多型性 DNA マーカーとは

多型性 DNA マーカーは大きく分けて以下のように分類できる (図 1)。

1. 1 塩基置換によるもの=RFLP

末梢白血球などから抽出した高分子量 DNA を制限酵素にて切断し、任意の DNA プローブとサザンハイブリダイゼーションしたときに出現する断片の長さの多型性 (=制限酵素断片長多型) である⁴⁾。点突然変異であるため対立遺伝子が 2 個である場合が多く、ある

個人がヘテロ接合体となる頻度 (ヘテロ接合性) が 0.5 前後と低くなる。最近、PCR で目的の領域を増幅して、PCR 産物の中の点突然変異の部位の有無を制限酵素の消化で確認する PCR-RFLP や、適当な制限酵素が発見できない場合には、ASO (allele specific oligonucleotide) を利用したドットプロット・ハイブリダイゼーションが利用されている。PCR を用いるこれら方法は、サザンハイブリダイゼーションのように大量の DNA を消費しなくてもよいのが利点である。

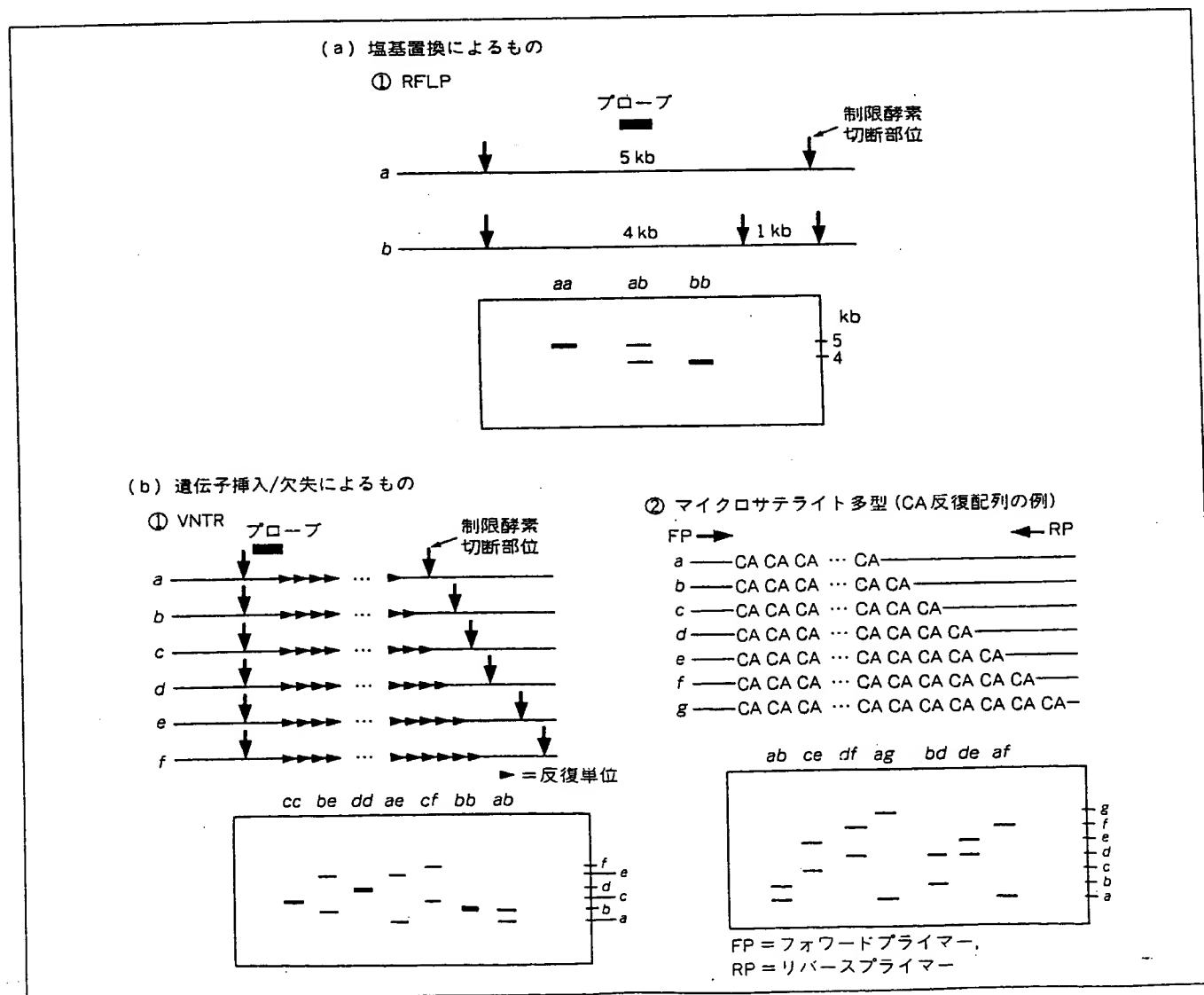


図 1 多型性 DNA マーカーの種類

2. 塩基の挿入/欠失によるもの

A. VNTRミニサテライト多型 (minisatellite)

1 単位が7~40 塩基対(bp)からなる短い単位の反復数に変異があるため生じる多型である⁵⁾。RFLPに比べ、対立遺伝子の数が多いため情報量が多いが、遺伝子座位が染色体上の末端側に分布し、偏っているのが欠点である。サザンハイブリダイゼーションで検出するのが一般的であるが、反復配列の前後をPCRで増幅して、PCR産物の長さの多型として検出できる例がある。たとえば、アポリポ蛋白質B遺伝子の3'末端や、インスリン遺伝子の5'末端のVNTRはPCRで簡単に増幅可能で、それぞれ高脂血症や糖尿病の病因遺伝子の解析に用いられている。

B. マイクロサテライト多型 (microsatellite)

1 単位が7bpまでの単位が反復する多型で、対立遺伝子の数が多く、ヘテロ接合性が80~90%となるマーカーが数多く報告されている。(CA)などのジヌクレオチドの反復配列は、ゲノム中に5~10万コピー存在し多型性に富む⁶⁾。PCRを利用するため、試料として用いる高分子量DNAは少量でよい。マイクロサテライトマーカーは、反復配列は、モノヌクレオチド(AAA...など)、ジヌクレオチド(CA, TA, CGなど)、トリヌクレオチド(CTG, CGGなど)、テトラヌクレオチド(TAAAなど)、ペンタヌクレオチドなどの種類があるが、CA反復配列は数が多いため、最もよく使用されている。ジヌクレオチドを中心とした各マーカーの情報はGenethon(Webは、<http://www.genethon.fr>)からインターネットで得られる。おもな情報としては、各マーカーのENBL/GenBankへのアクセス番号、男女平均と男女別にした近接するマーカーとの遺伝子距離、GDB登録番号、対立遺伝子の長さの数、ヘテロ接合性、プライマーの塩基配列、CEPH家系の#134702のヒトのバンドの長さ、最後にPCRの反応条件が記述されている。トリヌクレオチド以上のマイクロサテライト多型は、CHLC(Cooperative Human Linkage Center, Webは<http://www.chlc.org>)に登録されている。ある染色体の特定の領域の遺伝子地図を知りたい場合は、GDB(Genome Data Base, Webは、<http://gdbwww.gdb.org>)が役にたつ。特定の遺伝病の原因遺伝子の解析にどのような多型性DNAマーカーが必要であるかの情報は、OMIM(Online Mendelian Inheritance In Man, WebはGDBからNCBIの

中へ最近移動した。<http://www.ncbi.nlm.nih.gov>)が役だつ。マーカーの種類により日本人と白人の間で対立遺伝子頻度が大きく異なる場合がある⁷⁾。情報のほとんどは白人の解析によるため、日本人の解析ではほとんど役立たない場合もあり注意を要する。また、マイクロサテライト多型のプライマーは、米国の企業であるResearch Genetics(<http://www.resgen.com>)より販売されている。インターネットの画面上で、希望する染色体座位、ヘテロ接合性などを選んでから、必要なマーカーを直接注文することにより入手できる。

II. 多型性DNAマーカーの単離

重要な遺伝子から遺伝子多型を見つけた方法、直接塩基配列を決定する方法、SSCPで遺伝子変異を見だし、次に塩基配列を決定する方法、PCRである長さの遺伝子産物を作製し適当な種類の制限酵素で切断して多型を検出する方法などがある。塩基配列の多型が制限酵素で切断される場合は、PCR-RFLPを用いる。制限酵素で切断されない場合は、ASOや対立遺伝子に特異的なプライマーを用いて対立遺伝子を区別する。ある特定の遺伝子領域からマイクロサテライトを単離する方法は、YACクローンであれば、まずコスミドへサブクローニングする。作製したコスミドライブラリーをCAのオリゴヌクレオチドをプローブにしてハイブリダイゼーションを行ない、CA反復配列を含むコスミドを単離する。このコスミドを、*Sau3AI*などの制限酵素で消化し、プラスミドにサブクローニングして、再びCAのオリゴヌクレオチドでスクリーニングし、CAを含むクローンを単離する。このクローンのインサートの塩基配列を決定し、CA反復配列を挟んだ適当なプライマーを作製し、健康人6人ぐらいを解析して多型が出現するかどうか確認する。この方法の欠点は、プラスミドのインサートの長さが長い場合はCA反復配列を挟んだ塩基配列が決定できない、逆にインサートの長さが短い場合、適当な長さのプライマーの塩基配列が決定できないことなどである。CA反復配列のくり返し数が14以上のものは、多型性を示す確率が高い。林らは、コスミドから直接にCA反復配列の前後の塩基配列を決定する方法として、6種類のシーケンスプライマーを混合させる方法を開発している⁸⁾。

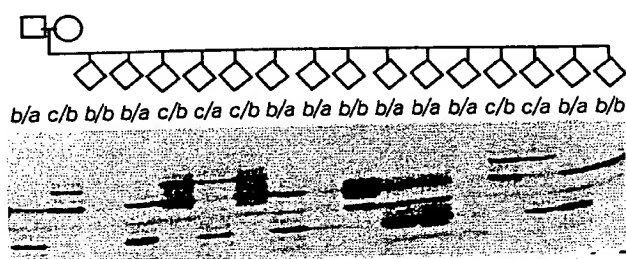
III. マイクロサテライト多型の検出方法

マイクロサテライト多型の検出には、一般的には、片方のプライマーを ^{32}P による RI 標識か、FITC やローダミンなどによる蛍光標識を行ない、塩基配列決定用の電気泳動で、遺伝子型を決定していく。図 2 には、ローダミン標識のプライマーを使用した例を示す。この STS は、ヒト染色体を顕微切断したゲノム由来のライブラリーから単離したマイクロサテライト多型で、CEPH 家系をタイピングした例である。GDB によって、D7S1682 (MS8-148) と命名され、国際的に登録されている⁹⁾。

IV. リンケージマッピング¹⁰⁾

1. 逐次検定法

図 3 に示すように、生殖細胞での減数分裂時、父方由来と母方由来の 2 本の相同染色体間の相同部分で交叉 (= 乗換え; crossing over) が生じ、結果として遺



115bp=a, 119bp=b, 121bp=c

図 2 CEPH 家系 (#1423) を利用した D7S1682 の遺伝子型決定 (FMBIO 使用)

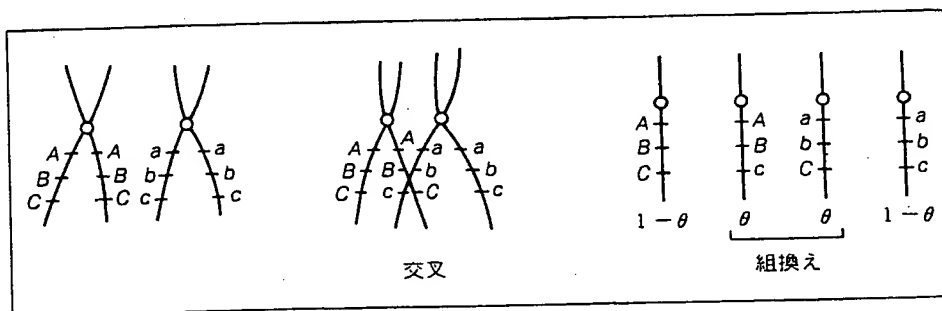


図 3 減数分裂時の交叉と遺伝子組換え

伝子の組換えが生じる。1 回の減数分裂につき、全染色体で約 50 カ所の交叉が生じる。遺伝子座 A と B のように 2 つの遺伝子が近接して存在する場合は、交叉の頻度は低くなる。ところが、遺伝子座 A と C のように、同一染色体上の 2 つの遺伝子距離がある程度離れると、交叉の頻度が高くなる。組換えの頻度は組換え率 $= \theta$ で表わされる。2 つの遺伝子座の距離が離れるほど組換え率 (θ) は大きくなり、逆に近くに座位する場合は連鎖していると表現し、組換え率はほとんど 0 となる。同じ染色体上の 2 つの遺伝子座がある程度離れると、座位間の交叉は 2 回以上となり奇数回と偶数回の頻度は等しくなるため、組換え率は $0.5 = 1/2$ となる。また、2 つの遺伝子座が別々の染色体上に位置する場合も、同一の生殖細胞に分離される確率は 0.5 となり、組換え率は 0.5 となる。2 つの遺伝子座の地図距離 (map distance) $= X$ と組換え率 $= \theta$ の関係が連鎖分析の基本となる。同一染色体の 2 つの遺伝子座間に、1 回の交叉しか生じないような場合の遺伝子距離は $X = \theta$ と定義される。X は、cM (センチモルガン) という単位で表わす。ヒトの遺伝子地図は、ゲノム上に存在する遺伝子マーカーによって作製され、地図の長さは男子で約 27.0 M (モルガン) と推定されている。男子の場合、27 M = 2700 cM となる。ヒトのゲノムは、ハプロイドあたり 3×10^9 bp よりなるため、1 cM は平均 $3 \times 10^9 \text{ bp} / 2700 = 1100 \text{ kb}$ となる。ショウジョウバエやマウスでは、計画的な交配で遺伝子地図が作成されているが、ヒトの場合は家系を統計学的に処理する連鎖分析で連鎖地図が作成 (linkage mapping) される。

ある優性遺伝病について 2 世代の家系 (図 4) について連鎖分析を行なうとする。黒印は患者を、病因遺伝子座位は G が変異対立遺伝子、g は正常対立遺伝子を示す。この座位から少し離れた位置に連鎖を検定する

A, B 対立遺伝子座位がある と仮定する。患者である親の I-1 の相が不明であるので、2 相 (①相引、②相反) が考えられる。たとえば、①相引の場合は、病因遺伝子 G が A 対立遺伝子と同一染色体にある場合である。この場合、子供である II-4, 5 の遺伝子型は組換え型となる。逆に②相反

の場合はII-1, 2, 3が, 組換え型となる。

相引のときは

$$\left(\frac{1-\theta}{2}\right)^1 \left(\frac{1-\theta}{2}\right)^2 \left(\frac{\theta}{2}\right)^1 \left(\frac{\theta}{2}\right)^1 = \left(\frac{1-\theta}{2}\right)^3 \left(\frac{\theta}{2}\right)^2$$

また, 相反のときは

$$\left(\frac{\theta}{2}\right)^1 \left(\frac{\theta}{2}\right)^2 \left(\frac{1-\theta}{2}\right)^1 \left(\frac{1-\theta}{2}\right)^1 = \left(\frac{\theta}{2}\right)^3 \left(\frac{1-\theta}{2}\right)^2$$

表1 ロッド値を求めるために必要な配偶子とその遺伝子頻度 (図4に示した家系に基づく)

GgABの配偶子	GA	gB	GB	gA
頻度				
相引①の場合	$\frac{1-\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{\theta}{2}$	$\frac{\theta}{2}$
	非組換え型		組換え型	
相反②の場合	$\frac{\theta}{2}$	$\frac{\theta}{2}$	$\frac{1-\theta}{2}$	$\frac{1-\theta}{2}$
	組換え型		非組換え型	
①の相の数	1	2	1	1
②の相の数	1	2	1	1
一般式用	a	b	c	d

となる。表1に示すように, 罹患した親は, 相引か相反か不明であるので, それぞれの確率を足して1/2倍する。したがって, 組換え率 θ のとき, 病因遺伝子座位とAB遺伝子座位が連鎖しているとする, その確度=尤度 (likelihood) = L は,

$$P(f; \theta) = \frac{1}{2} \left\{ \left(\frac{1-\theta}{2} \right)^3 \left(\frac{\theta}{2} \right)^2 + \left(\frac{\theta}{2} \right)^3 \left(\frac{1-\theta}{2} \right)^2 \right\}$$

$$= \frac{1}{2^5} \{ (1-\theta)^3 \theta^2 + \theta^3 (1-\theta)^2 \}$$

となる。

2つの遺伝子間の組換え率(θ)を0.0から0.5まで変更したときの確率 $P = (f/\theta)$ と連鎖がない場合($\theta = 0.5$)の確率 $P = (f/0.5)$ との比の対数がロッド値 (lod score; Z) である。図4の家系のロッド値は, 表2のようになる。もし, 図4の家系が3世代の家系であり, I-1の遺伝子型においてGとAが同一染色体上に存在するとわかっていれば, II-4, 5は組換え型となる。

尤度は,

$$P(f; \theta) = \frac{1}{2^5} \{ \theta^2 (1-\theta)^3 \}$$

となり, ロッド値は表2に示すようになる。

組換え率に影響を与えるものとして, 年齢と性がある。大部分の遺伝病は出生時は浸透率 (penetrance = 発症している個体/病因遺伝子をもって個体) は0%であるが, ある年齢ではほとんど100%となる。たとえば, 筋緊張性ジストロフィー症の場合は, 原因遺伝子を親より受け継いだヒトの50%が30歳までに発症し, 老人になっても全員が発症せず浸透率は94%である。浸透率が90%であれば, 組換え率(θ)が0.01で完全浸透と同じロッド値を得るためには約2倍の観察数が必要となる。浸透率が50%であれば, 約4倍の観察数が必要となる。2つの遺伝子間の組換え率は,

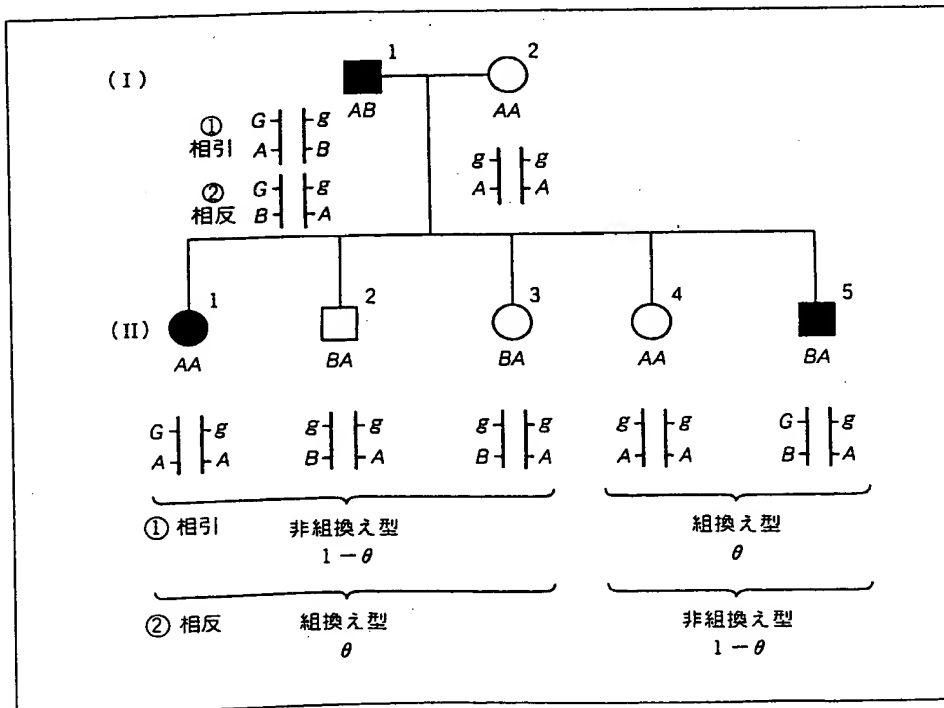


図4 常染色体性優性遺伝病の2世代の家系

■, ●は患者。G, gは責任遺伝子座位。G: 異常遺伝子, g: 正常遺伝子。A, Bは責任遺伝子に連鎖した遺伝子座位。

表 2

組換え率 (θ)	0	0.1	0.2	0.3	0.4	0.5
A 公算比	0	0.130	0.410	0.706	0.922	1
A ロッド値	$-\infty$	-0.887	-0.388	-0.151	-0.035	0
B 公算比	0	0.233	0.655	0.988	1.106	1
B ロッド値	$-\infty$	-0.632	-0.184	-0.005	0.044	0

ロッド値 $=\log_{10}$ [公算比], A: 2世代の家系, B: 3世代の家系, 図4のI-1においてGとAが同一染色体上に存在するとわかる。

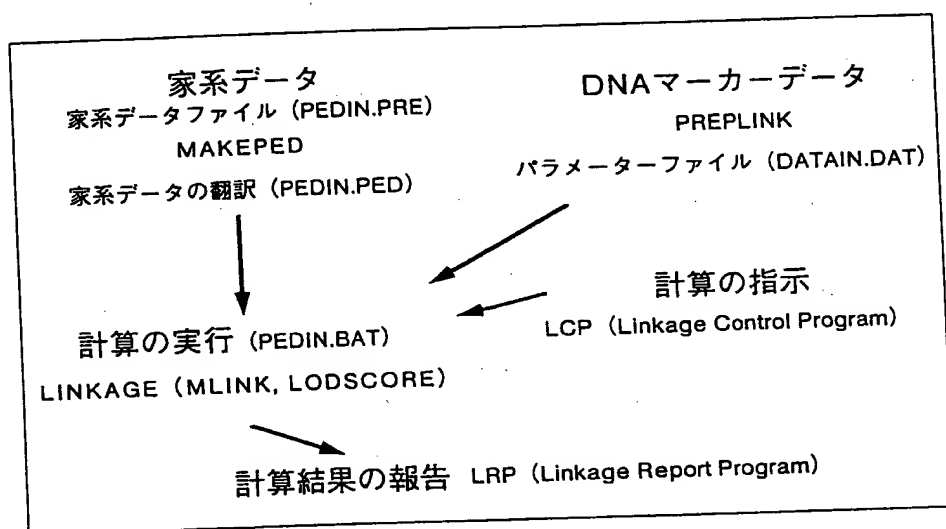


図 5 "LINKAGE" を作働させるとき必要な各プログラム (紙野による)

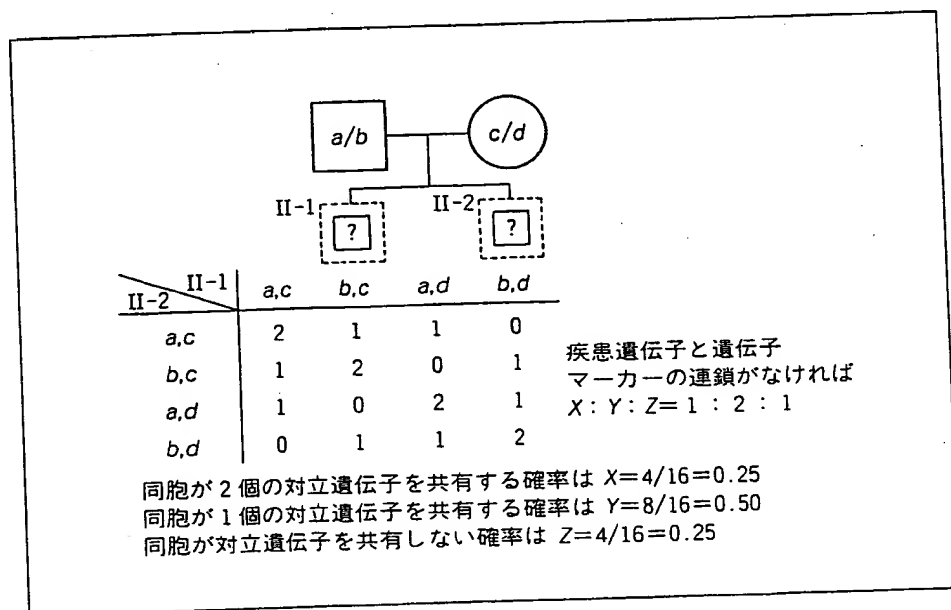


図 6 罹患同胞対法の原理

多くの場合女性のほうが男性より大きい。したがって、ある2つの遺伝子間で男性のみ連鎖が証明される場合がある。

ロッド値は各核家族ごと加算される。ロッド値が合計+3以上であれば両遺伝子の座位の間は連鎖があると判定し、-2以下であれば連鎖なし、-2から+3の間は判定が不可能でさらに多くのデータを必要とする。ヒトの全ゲノムで均等に 20 cM ($\theta=0.20$) ずつ分布した 150 個の DNA マーカーがあれば、理論上すべての遺伝病の原因遺伝子座位決定に対処できると考えられている。現在までに単離されている座位の判明した多型性マーカーは 5000 個以上あるので、大部分の遺伝病の座位は決定できる状態にある。実際のロッド値の算出には、手計算よりもコンピュータプログラムである "LIPED" あるいは、"LINKAGE" を使用する場合が多い。一般には、最大のロッド値 (Z_{\max}) を示す組換え率 (θ_{\max}) が2つの遺伝子の間の距離になる。たとえば、筋緊張性ジストロフィー症座位とアポリポ蛋白質 CII (APOC2) との間の連鎖分析で、 θ_{\max} (最大のロッド値をとるときの組換え率) $=0.03$ で、 $Z_{\max}=26.01$ である。 θ_{\max} の 95% 信頼限界が、0.015~0.075 である結果より、DM 遺伝子と APOC2 の遺伝子距離は、1500 kb から 7500 kb (1 cM を 1000 kb とする) の範囲

にある。図5に実際の”LINKAGE”で必要なパラメーターを示す。

遺伝子座位が判明しない多型性遺伝子マーカーの座位を決定する場合は、CEPH家系を用いた連鎖分析を行なう。すでに、5000個以上の遺伝子マーカーの座位が決定されているので、2点解析では、図4でのGg座位とAB座位の間の検定となる。

2. 罹患同胞対法 (affected sib pair method)

罹患同胞対法は、疾病遺伝子のマッピングに用いられる方法で、広義のリンクージマッピングになる。逐次検定法のような2~3世代にわたる大家系の構成員のサンプルを集める必要はなく、罹患した2人の同胞と両親のサンプルを集めるだけでよい。さらに、この分析法の利点は、疾患の遺伝形式(優性遺伝か劣性遺伝か)や浸透率(連鎖分析法はロッド値を算出する際、重要となる)を考慮に入れる必要のないことである。逆に、両親の4本の対立遺伝子を区別できる遺伝子マーカーを得る必要がある。この分析法に適切な遺伝子マーカーは対立遺伝子が多いHLA抗原やGm型などで、ABO血液型や点突然変異によるRFLPは対立遺伝子が2~3種であるためこの分析に用いてもあまり情報は得られなかった。図6に示すように、両親の4つの対立遺伝子がどれもヘテロ接合体(a, b) \times (c, d)である子供の遺伝子型は、(a, c), (b, c), (a, d), (b, d)のいずれかになる。罹患した2人の子供の遺伝子型は16通りの組合せがある。2人が2個の同じ遺伝子を共有する確率は $x=4/16=0.25$ となり、1個の遺伝子を共有する確率は $Y=8/16=0.5$ 、まったく遺伝子を共有しない確率は $Z=4/16=0.25$ となる。原因遺伝子と遺伝子マーカーが、密に連鎖していなければ遺伝子型の分布に偏りはなく $X:Y:Z=1:2:1$ となるが、もし、連鎖した遺伝子マーカーであれば偏りが生じる。この偏りの程度は、遺伝形式と原因遺伝子の遺伝子頻度によって異なる。病気の主効果遺伝子の頻度から、劣性遺伝と優性遺伝に分けた X, Y, Z の偏りを、期待される頻度として計算する。実際に、自己免疫疾患であるバセドウ病の2つの原因遺伝子座位、HLA座位とGm座位は本法で決定され、いずれも劣性遺伝であった。ま

た、同時に原因遺伝子の遺伝子頻度も推定されたが、これは疫学調査とまったく一致していた。この方法を用いた成人病の解析が始まっている。マイクロサテライト多型の出現は、罹患同胞対法の解析を飛躍的に発展させた。実際の応用として、英国ではインスリン依存性糖尿病(IDDM)の家系を102家系集めて、原因遺伝子座位を決定する作業が始まっている。

おわりに DNAマーカーは、PCRによるマイクロサテライトマーカーの出現で大きく変化した。ゲノム計画も飛躍的に進歩した。まず、マイクロサテライトマーカーを用いた連鎖分析や関連研究で、染色体上の遺伝子座位を決定することを第一段階とする。次に、発現しているのが確認され、さらに遺伝子座位も決定されたESTs (expressed sequence tags) のなかから、責任遺伝子そのものを単離同定する作業、いわゆるpositional candidateアプローチが主流となるであろう。

文 献

- 1) 三木哲郎・荻原俊男：日本臨床1994年特別号「臨床分子生物学」, 488-495 (1994)
- 2) Dib, C. *et al.* : *Nature*, 380, 152-154 (1996)
- 3) Hudson, T. J. *et al.* : *Science*, 270, 1945-1954 (1995)
- 4) Botstein, D., White, R. L., Skolnick, M., Davies, R. W. : *Am. J. Hum. Genet.*, 32, 314-331 (1980)
- 5) Nakamura, Y., Leppert, M. E., O'Connell, P. O., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., White, R. L. : *Science*, 235, 1616-1622 (1987)
- 6) Weber, J. L., May, P. E. : *Am. J. Hum. Genet.*, 44, 388-396 (1989)
- 7) Kamino, K., Nakura, J., Kihara, K., Ye, L., Nagano, K., Ohta, T., Jinno, Y., Niikawa, N., Miki, T., Ogihara, T. : *Hum. Mol. Genet.*, 14, 407-419 (1993)
- 8) Iizuka, M., Sekiya, T., Hayashi, K. : *Technique*, 3, 171-173 (1991)
- 9) Nakura, J., Ye, L., Miki, T., Mitsuda, N., Ogihara, T. : *Jpn. J. Hum. Genet.*, 39, 447-449 (1994)
- 10) 安田徳一：日本医師会雑誌, 89, 563-572 (1983)